U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 80

Toward an Operational Four-Dimensional Data Assimilation System

Ronald D. McPherson
and
William D. Bonner
Development Division

NOVEMBER 1972

I. Introduction

The National Meteorological Center has operated for many years on a twice-daily cycle of data collection, processing, and analysis centered on 00 GMT and 12 GMT. This cycle accepts all data judged to be useful from observations taken at or near those times. Until recently, observations taken at other times were not used. With the advent of satellite-borne meteorological sensors, however, new data became available which considerably improved the coverage over previously data-sparse areas. These data are not taken at the standard synoptic times. It thus became necessary to devise ways of assimilating these new data in order that they have an impact on operational numerical weather prediction.

One step has been taken to allow a limited but important species of asynoptic data to enter the NMC data base; this step will be described in the next section of this report. However, it is generally conceded that this one step is inadequate and more sophisticated systems must be developed in order to permit maximum utilization of the data. This report outlines a proposed evolutionary approach to a four-dimensional analysis system capable of being implemented operationally.

In recent years, interest in the data assimilation problem has been very strong, and the literature reflects this interest. Most of the articles in the literature have dealt with what might be called the dynamic assimilation method. In this approach, the basic vehicle is a prediction model; experiments have been done both with primitive equation models (e.g., Morel, et al., 1971; Jastrow and Halem, 1971) and with quasi-geostrophic models (Bengtsson and Gustavsson, 1972). The general approach is as follows: the model is integrated forward in time, and as the integration reaches the time at which an observation is made (an indirect sounding from a satellite, for example), the observation replaces (and presumably corrects) the forecast value at that point in space and time. This process is usually described as direct, continuous insertion. Variations have been proposed; one is indirect insertion, in which a local objective analysis is performed about the point of insertion, thus recognizing that an observation influences the field in its neighborhood rather than at one point only. Another variation is called intermittent insertion, in which data observed during a given interval is considered valid at a particular time, and is inserted en masse into the model at that time. This process usually results in a "shock" to the numerical model which is manifested as gravitational oscillations. Various devices have been employed to damp these oscillations; Morel and his colleagues (Sadourny, 1972) have used a special viscosity which affects only the divergence. Hayden (1972), in one of the few real-data experiments, used the Euler-backward time integration scheme in forward-backward cycling about the time of insertion to damp the shocks. Extensive cycling in this manner might prove to be costly for operational use.

A second approach employs the calculus of variations to obtain analyses which obey certain dynamical constraints. Sasaki (1971) has been a proponent of this method for more than a decade. As yet, results which would clearly indicate its applicability to the global problem have not appeared in the literature.

There is a third approach which is essentially statistical in nature, and which is an extension of Gandin's (1963) method of optimum interpolation to four dimensions. Peterson (1968) has described this method. More recently, Rutherford (1972) in Canada has published a paper suggesting the method's applicability to the global problem.

A fourth approach might be called "spectral." Only one paper on the subject has appeared thus far, published by Dixon, et al., (1972). Their idea is to represent meteorological variables by series of polynomials, and to fit the observations to these polynomials by least-squares. The claim is that the method can be extended to the time dimension. Another approach which has not yet been published is that of Flattery, wherein his Hough-function global analysis scheme would be combined with a spectral forecast model. Such a model exists, but development work is not complete.

It would be of some interest to investigate each of these methods. However, the resources available are not adequate to complete this task within the time limits imposed by forthcoming operational requirements. Instead, it is necessary to focus on one method which seems most promising for operational implementation.

The dynamic method has been experimented with more extensively than any other, and appears to be favored by most investigators. It is noteworthy, however, that most of the experiments have been conducted with simulated, rather than real data, and results from the few real-data experiments (e.g., Hayden, 1972) are far from encouraging. It therefore seems unwise to proceed immediately to an operational direct, continuous dynamic assimilation system. We have instead chosen to proceed in the direction of such a system through an evolutionary process, consisting of several well-defined stages.

The next section describes briefly the current state of four-dimensional assimilation at NMC. It is followed by a section outlining the next stage, which is presently intended for implementation in the spring of 1973. The fourth section describes the assimilation system envisaged for use during 1974, and the final section presents a summary of the evolutionary process and speculation concerning post-1974 systems.

II. Stage I. The current state of four-dimensional assimilation at NMC

At the present time, the operational analysis/forecast system consists of two main elements: the operational objective analysis system, which is a successive-approximation method devised by Cressman (1957), and a six-layer primitive equation prediction model (6L PE) designed by Shuman and Hovermale (1968). The domain of operation is a polar-stereographic projection of the Northern Hemisphere. An operational cycle consists of collecting all 00 GMT or 12 GMT data from the Northern Hemisphere down to about 10N; data that reach NMC prior to about 3.5 hours after observation time are entered into the data base. The operational analysis then uses this data and a previous 12-hour forecast as a first guess, and produces the objective analyses of meteorological fields that are necessary to specify the initial state of the atmosphere for the prediction model. Once this is done the prediction model is integrated out to 48 hours; the 12-hour forecasts from 00 GMT data are saved to provide the first guess for the next operational analysis.

Once each day, a delayed, or "final", analysis is done. This re-analysis uses 12 GMT data that reach NMC prior to ten hours after observation time. From the final analysis, a 12-hour forecast is made to serve as a first guess for the next 00 GMT operational analysis.

Until approximately two years ago, only synoptic data (including aircraft reports within ±3 hours of synoptic time) from 00 GMT and 12 GMT were accepted into the data base. With the operational availability of SIRS-B temperature soundings on an orbit-by-orbit basis, an effort was made to incorporate those asynoptic data into the data base. This effort was largely the work of A. J. Desmarais, who has recently documented the work in an NMC Technical Memorandum (Desmarais, 1972); his technique constitutes what is called Data Assimilation Stage I in this report. It is illustrated in Figure 1.

Briefly, soundings retrieved from satellite radiance measurements at asynoptic times are "updated" or "downdated" to the nearest synoptic time by applying to the sounding a correction based on the forecast tendency of the parameter, either temperature or height. Soundings retrieved during the six-hour period prior to synoptic time are updated to synoptic time; those retrieved up to six hours later are downdated. The forecast tendency is determined as a linear trend between the 6- and 18-hour operational forecasts. These forecasts are also used as the first guess for the temperature retrievals from observed radiances; this is, of course, for the Northern Hemisphere down to about 18N, only. In the Tropics, a standard sounding is employed, and a temperature analysis based on persistence and current conventional data is used in the Southern Hemisphere. This procedure is employed primarily because forecasts for those regions are not currently available.

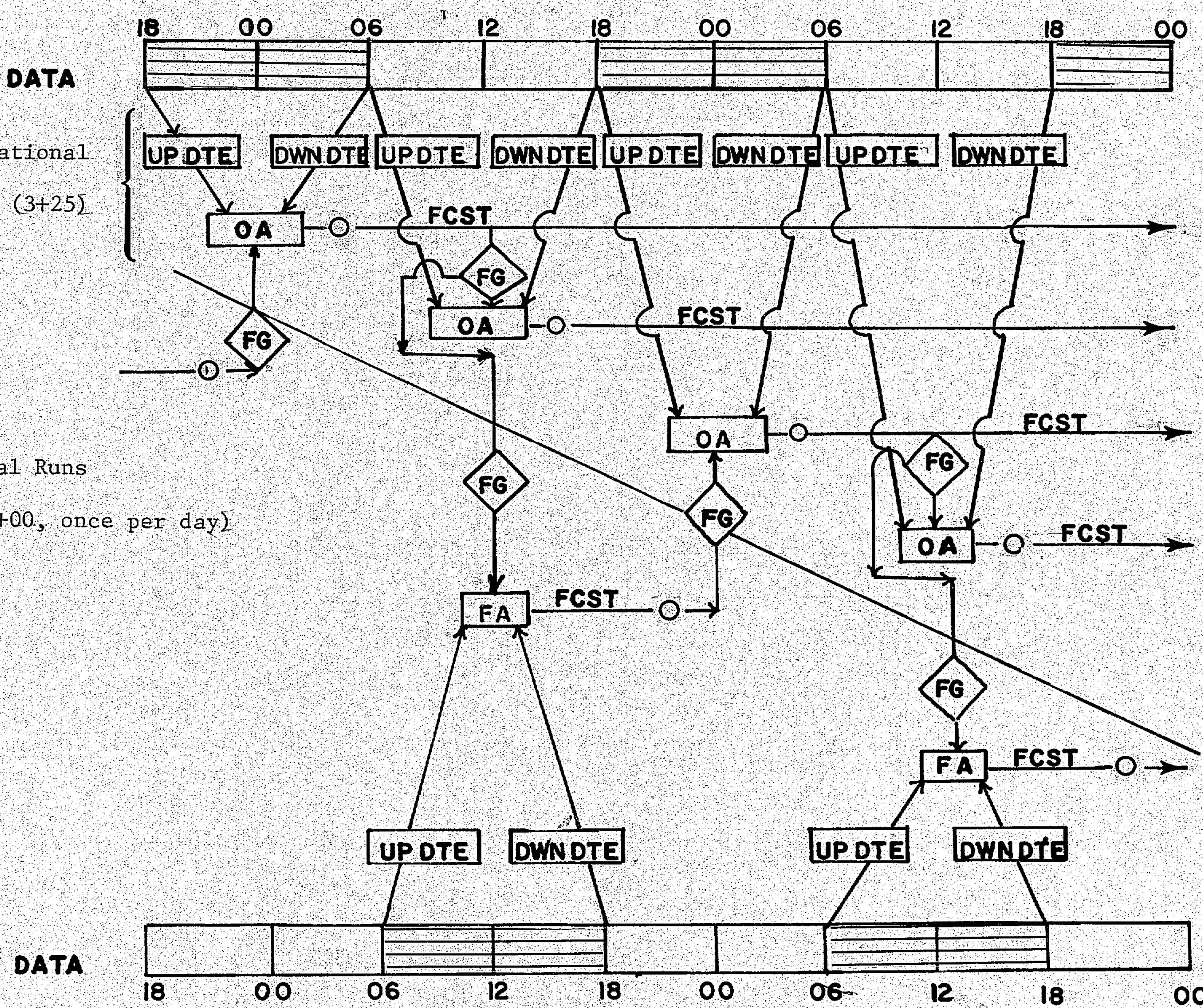
STAGE I

ASSIMILATION SCHEME

(Current - April 1, 1973)

Operational Runs (3+25)

Final Runs (10+00, once per day)



OA = Operational analysis

FA = Final analysis

FG = First guess for OA or FA

Update/downdate uses 6-18 hr operational forecast tendencies to correct data to synoptic time.

Operational model is the 6L PE, twice daily at 3+25; once daily to 12-hours from 12 GMT final analysis.

Figure 1

STAGE I

RETRIEVAL SYSTEM

Northern Hemisphere: Same as Stage II.

Tropics: First guess is a standard tropical sounding, used between 18N and 18S.

Southern Hemisphere: First guess is a current temperature analysis, based on persistence and conventional data.

In Stage I, then, asynoptic satellite data go into the data base by means of a temporal correction of the observation to synoptic time. As indicated in Figure 1, this is done twice daily at operational time (3+25), and repeated once daily at final time (10+00). Aircraft reports within ±3 hours of 00 GMT or 12 GMT enter the data base as synoptic data, without updating. No other types of asynoptic data, or synoptic data at times other than 00 GMT and 12 GMT are considered. Such data are therefore not used in operational numerical weather prediction.

III. Stage II

For Stage II, we wish to take steps in the direction of continuous assimilation, and simultaneously attempt to remedy the most outstanding deficiencies of Stage I. Specifically, it is desired to:

(1) replace the current analysis system with a global system;

(2) arrange the insertion of satellite data so that updating is not necessary;

(3) accept other types of previously available but unused synoptic data, such 06 GMT and 18 GMT surface reports and radiosondes;

(4) provide an improved first guess for satellite retrievals in the Tropics and Southern Hemisphere.

The proposed Stage II assimilation system is illustrated in Figure 2a; Figure 2b shows schematically the retrieval system.

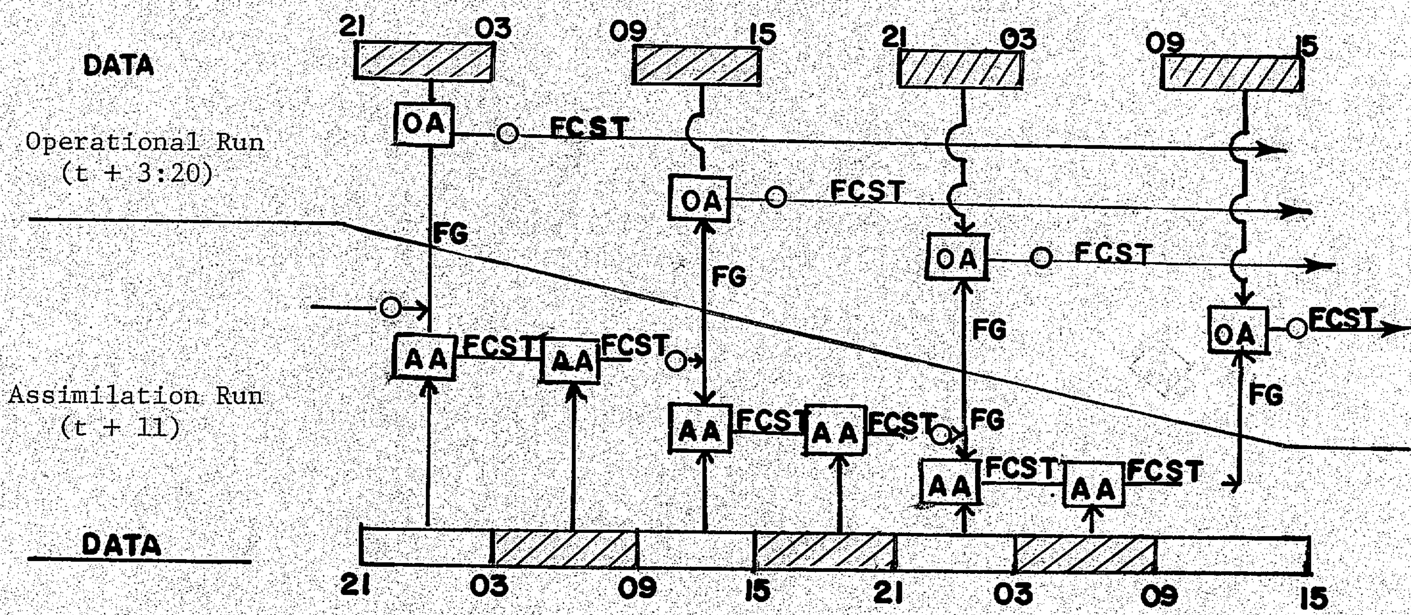
The fundamental changes involved in the transition from Stage I to Stage II are the following:

(1) the Cressman successive-approximation objective analysis scheme is supplanted by Flattery's (1971) spectral global analysis as the operational analysis system;

(2) the once-daily final analysis and 12-hour forecast at 22 GMT (10+00) is replaced by a twice-daily assimilation run at 11+00 consisting of a final spectral global analysis using all available synoptic data, followed by a 6-hour forecast using the Stackpole (1972) eight-layer primitive equation model (8L GLOBAL) on a global 5-degree mesh, a global re-analysis with insertion of the available additional data, and finally another 6-hour forecast to serve as a first guess for the next operational analysis.

STAGE 2

ASSIMILATION SCHEME (April 1, 1973)



OA is operational analysis (Flattery)

AA is assimilation analysis (Flattery)

FG is first guess for OA or AA

TIMES are data times

CIRCLES are start times (real time) for forecast runs.

DATA initially means VTPR only. May extend later to mean any data including 06 and 18 GMT surface reports.

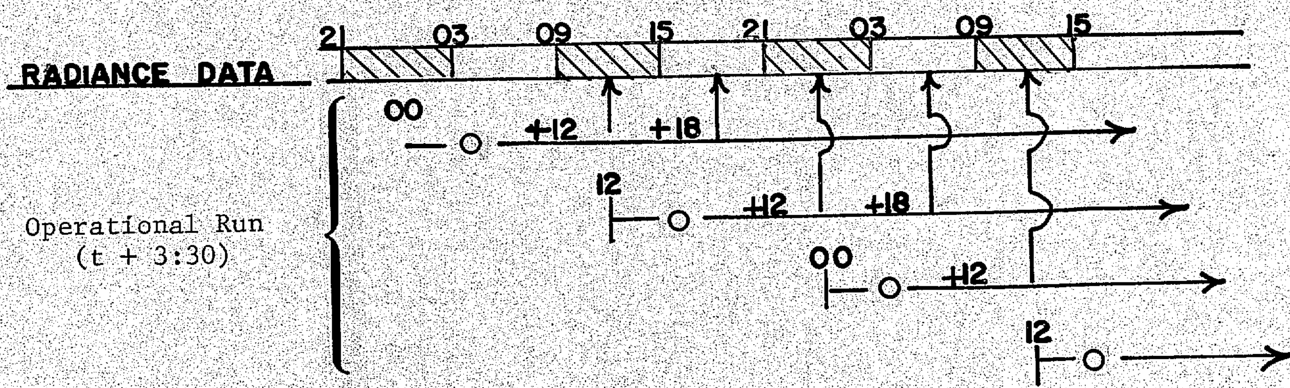
THE OPERATIONAL MODEL is the 6L PE

THE ASSIMILATION MODEL is the 8L GLOBAL with 5 deg grid.

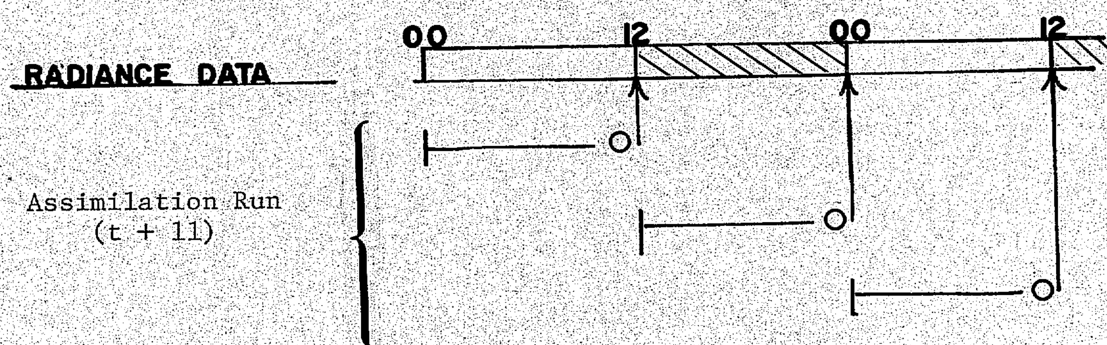
Figure 2a

STAGE 2

RETRIEVAL SYSTEM - NORTHERN HEMISPHERE



RETRIEVAL SYSTEM - SOUTHERN HEMISPHERE AND TROPICS



Retrievals in N. Hemisphere (N. of 18N) - Use 12 and 18 hr forecast from operational runs of 6L PE. Everywhere else, retrievals use 12 hr forecasts from assimilation runs of 8L GLOBAL model.

Figure 2b

Retrieved soundings from satellite-borne sensors are accepted in 6-hour blocks; that is, observations within ±3 hours of synoptic time are considered as synoptic data. As indicated in Figure 2a, satellite data observed between 21 GMT and 03 GMT are treated as synoptic data valid at 00 GMT and entered into the operational analysis for that time. Similarly, satellite data observed between 09 GMT and 15 GMT enter the operational analysis for 12 GMT as synoptic data. Therefore, updating or downdating of the satellite observations is omitted.

In the assimilation runs, begun at 11 hours after observation time, all conventional synoptic data collected up to the 11+00 cutoff time, as well as the satellite data within ±3 hours of synoptic time, are entered into the final global analysis. For example, for the 00 GMT synoptic time, the assimilation run begins at 11 GMT with a final analysis valid at 00 GMT. All conventional synoptic data for 00 GMT, plus the satellite observations between 21 GMT and 03 GMT, are available for the final analysis. From the final analysis, a 6-hour forecast valid at 06 GMT is made. This forecast serves as a first guess for a global analysis, which accepts satellite data observed in the 6-hour block between 03 GMT and 09 GMT, treated as valid at 06 GMT, plus any conventional 06 GMT synoptic data. Another 6-hour forecast is then made, using the 06 GMT reanalysis as initial data, to provide a forecast valid at 12 GMT. This forecast serves as the first guess for the operational analysis at 12 GMT. The assimilation run for 12 GMT begins at 23 GMT, and similarly consists of two 6-hour forecasts with a reanalysis valid at 18 GMT into which satellite data between 15 GMT and 21 GMT plus conventional 18 GMT data are inserted.

The equivalent of twice-daily 12-hour global forecasts would thus be made, with provision for inserting satellite data in 6-hour blocks without updating. Conventional synoptic data for the 06 GMT and 18 GMT intermediate times would also be utilized. The first-guess forecast provided for retrievals north of 18N would remain the 6- to 18-hour forecasts from the operational 6L PE. This does not represent a change from Stage I, except that the analysis from which the 6L PE runs uses an 8L GLOBAL forecast as a first guess. However, 12-hour forecasts for the area south of 18N would be available twice daily for retrieval purposes. Although these would be available on a delayed basis, so that their use as a first guess for retrievals would be essentially 12-hour persistence, we feel that this would represent an improvement over the present system.

IV. Stage III

Stage II represents a step toward an operational global analysis/forecast system with continuous four-dimensional data assimilation. The operational analysis is global, and the assimilation runs are conducted with a global analysis/forecast system. The technique of data assimilation might be characterized as delayed (accomplished in a "final" run), indirect (data is inserted by means of an analysis procedure, rather than directly at a grid point), and intermittent (insertion is at 6-hour intervals).

Additional steps are proposed for Stage III. The major changes are:

(1) the operational 6-layer hemispheric prediction model is replaced by the 8L GLOBAL model on a 2.5 degree mesh;

(2) the spatial resolution of the assimilation model is similarly increased to 2.5 degrees, and the frequency of reanalysis and data insertion is increased to 3 hours;

(3) forecasts for retrieval first-guesses are available globally from the operational model.

It is assumed that in Stage III, twice-daily global forecasts to at least 48 hours will be made routinely using the eight-layer primitive equation model on a 2.5 degree latitude-longitude mesh. The 9- to 18-hour period of these forecasts will be made available for use as the first guess for retrievals; this is illustrated in Figure 3b.

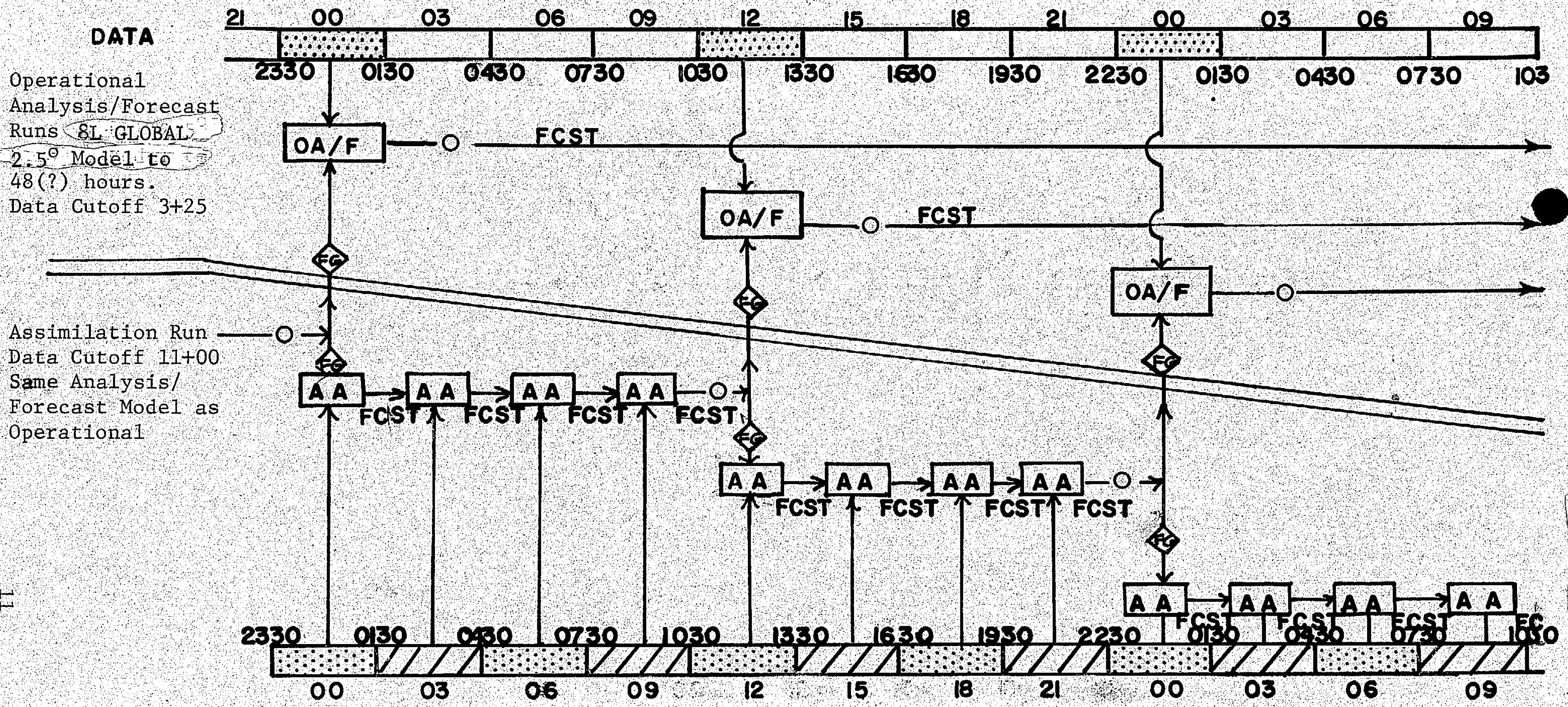
Figure 3a shows the Stage III operational and assimilation system. Aside from the major change in prediction models, the only change envisioned for the operational run is that the interval in which satellite data is to be considered as synoptic is tightened to ±1.5 hours. Thus, at 00 GMT, only data from orbits between 2230 GMT and 0130 GMT are treated as synoptic data. Likewise, data from orbits between 1030 GMT and 1330 GMT are considered as synoptic at 12 GMT. All other satellite data enter the data base through the twice-daily assimilation runs.

The major change in the assimilation runs, aside from increasing the spatial resolution of the model to 2.5 degrees, is that each run consists of four 3-hour forecasts with reanalysis at intermediate times, rather than two 6-hour forecasts. This represents another step toward continuous assimilation. It remains intermittent, but as the frequency of data insertion increases, the distinction between continuous and intermittent assimilation is blurred. The indirect and delayed character of the Stage II scheme is retained in Stage III.

It should be noted that Stage III depends on the acquisition of a more powerful computer, whereas Stage II does not.

V. Summary of the Evolutionary Process and Post-Stage III Speculation

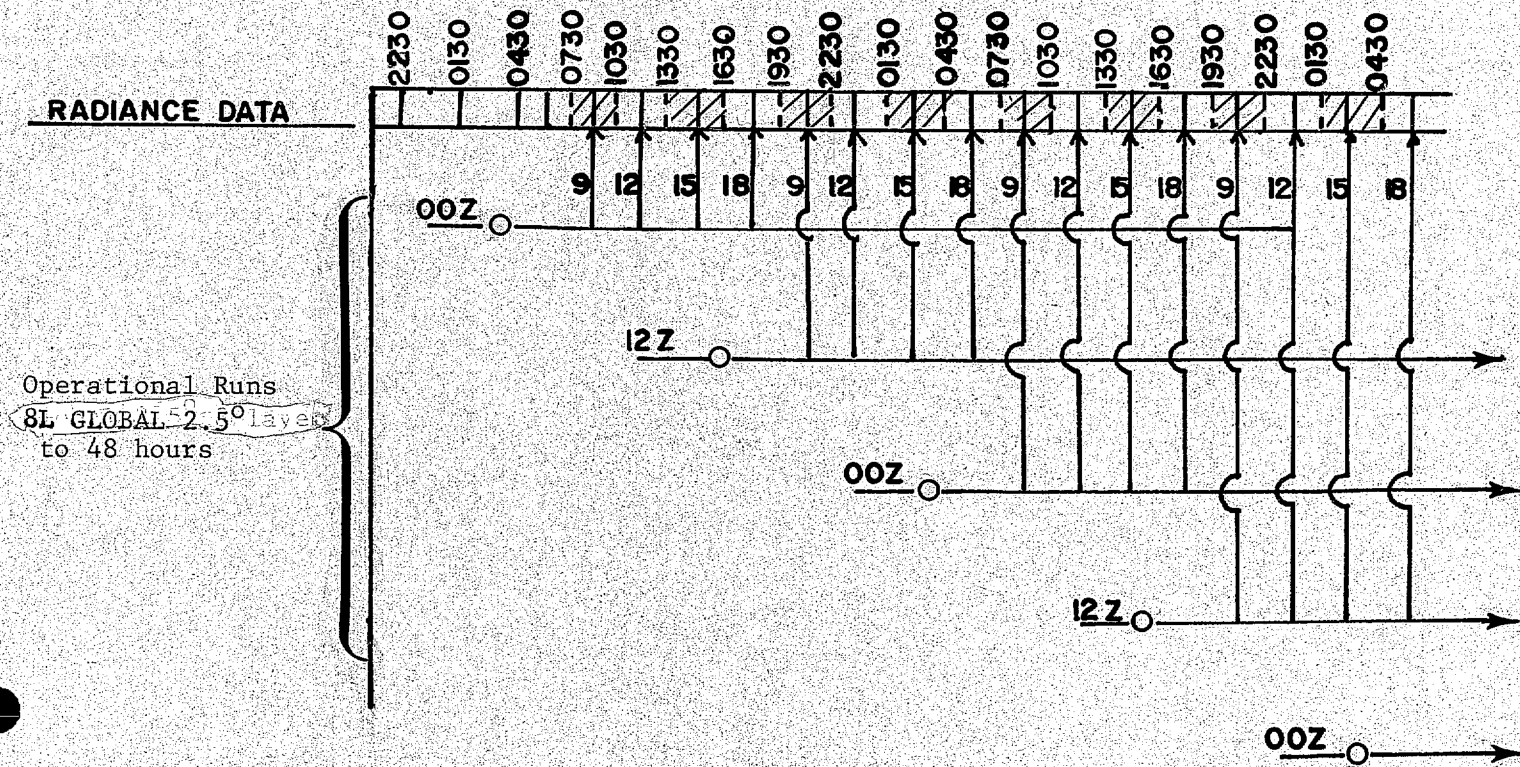
A summary of the characteristics of the first three stages, and the major changes involved between stages, is given in Table I. We have consciously rejected the temptation to plunge into direct, continuous assimilation; rather, we have proposed a process which moves in that direction from a more-or-less known starting point in a series of deliberate



OA/F = Operational analysis (Flattery)/forecast (Stackpole) system
AA = Assimilation analysis (Flattery)
FG = First guess for OA or AA
TIMES refer to the time of observation of the data in GMT
DATA refers to both satellite and conventional data sources at both synoptic and asynoptic times
[illegible]
CIRCLES indicate real time of data cutoff and beginning of either operational or assimilation run.

Figure 3a

RETRIEVAL SYSTEM - STAGE III



Times across the top are times of observation of each block of radiance data.
Times across the second row indicate the length of each forecast used as a first guess for retrievals.

Circles indicate real time of beginning of operational analysis/forecast cycle, assuming a 3+25 data cutoff.

Figure 3b

Table 1

SUMMARY OF CHARACTERISTICS OF THE FIRST THREE STAGES OF OPERATIONAL FOUR-DIMENSIONAL DATA ASSIMILATION

Stage/Date	Analysis	Model	Retrieval First Guess	Assimilation Method	Major Changes
I. Current-1 Apr. 73	Cressman successive approximation method over Northern Hemisphere octagon. Twice daily at 3+25. One final analysis at 10+00.	Operational: Shuman 6L PE mesh length 381 km at 60N, over Northern Hemisphere rectangle.	North of 18N Uses operational 12-18 hr forecasts. Tropics Standard sounding. Southern Hemisphere Temperature analysis based on conventional data and persistence.	Updating 6-18 hr forecast tendencies from operational model.	—
II. 1 Apr. 73-1 Jan. 74.	Flattery global spectral method. Twice daily. Operational at 3+25. Assimilation at 11+00.	Operational: Same as Stage I. Assimilation: 8L GLOBAL, 5° mesh.	North of 18N. Same as Stage I. Elsewhere 12-hr forecast from assimilation run valid at beginning of retrieval period.	Twice daily assimilation runs with 8L GLOBAL on 5° mesh. Data insertion and reanalysis each 6 hours.	1. Spectral global analysis replaces hemispheric analysis as operational system. 2. No updating. One final analysis run replaced by two assimilation runs. 3. Improved first guess for retrievals in the Tropics and Southern Hemisphere.
III. 1 Jan. 74-	Same as Stage II.	Operational and assimilation: 8L GLOBAL, 2.5° mesh.	9-18 hr global forecasts from operational model.	Same as Stage II except prediction model has 2.5° mesh length, data inserted and reanalyzed each 3 hours.	1. 8L GLOBAL on 2.5° mesh replaces 6L PE model as operational system. 2. Assimilation run data insertion frequency increased to 3 hours, and resolution of assimilation model increased to 2.5°. 3. Satellite retrievals given global forecasts at high resolution for first guess.

steps. We believe that this approach, which gives a series of benchmarks against which subsequent modifications may be compared, is a sound one for an operationally-oriented organization such as NMC to follow, and that it stands quite a reasonable change to succeed.

Beyond Stage III, we can only speculate as to the character of the operational data assimilation system. It may be desirable and economically feasible to further increase the spatial resolution of the prediction model. A further increase in the frequency of insertion may also be desirable; another doubling of the insertion frequency would have data being inserted in 1.5 hour blocks, which is almost orbit-by-orbit. It may also prove desirable to move away from delayed (i.e., at final time) assimilation toward real-time assimilation with direct feedback between retrieval and assimilation. Justification for these rather expensive improvements will arise only through the experience gained in Stage III.

REFERENCES

Bengtsson, L., and N. Gustavsson, 1972: "Assimilation of Non-Synoptic Data," Swedish Meteorological and Hydrological Institute manuscript, to be published.

Cressman, G., 1957: "An Operational Weather Analysis System," Monthly Weather Review, Vol. 87, No. 10.

Desmarais, A., 1972: "Updating Asynoptic Data for Use in Objective Analysis," NOAA Tech. Memo. NWS NMC-51.

Dixon, R., E. A. Spackman, J. Jones, and Anne Francis, 1972: "The Global Analysis of Meteorological Data Using Orthogonal Polynomial Base Functions," Journal of the Atmospheric Sciences, Vol. 29, No. 4.

Flattery, T., 1971: Untitled and unpublished manuscript made available as a personal communication.

Gandin, L. S., 1963: Objective Analysis of Meteorological Fields. Translated from Russian, TT 65-5007, available from National Technical Information Service, Commerce Department, Springfield, Va.

Hayden, C. M., 1972: "Experiments in the Four-Dimensional Assimilation of Nimbus 4 SIRS Data," submitted for publication in Journal of Applied Meteorology.

Jastrow, R., and M. Halem, 1971: "Asynoptic Assimilation of Satellite Temperature Data," presented at the International Symposium on Four-Dimensional Data Assimilation for GARP, 19–22 April 1971, Princeton, N.J.

Morel, P., G. Lefevre, and G. Rabreau, 1971: "On Initialization and Non-Synoptic Data Assimilation," Tellus, Vol. 23, No. 3.

Peterson, D. P., 1968: "On the Concept and Implementation of Sequential Analysis for Linear Random Fields," Tellus, Vol. 20, No. 7.

Rutherford, I., 1972: "Data Assimilation by Statistical Interpolation of Forecast Error Fields," Journal of the Atmospheric Sciences, Vol. 29, No. 5.

Sadourny, R., 1972: "Forced Geostrophic Adjustment in Large Scale Flow," Laboratoire de Météorologie Dynamique du C.N.R.S., (University of Paris) manuscript.

Sasaki, Y., 1970: "Numerical Variational Analysis Formulated Under Constraints as Determined by Long-Wave Equations and a Low Pass Filter," Monthly Weather Review, Vol. 98, No. 10.

Shuman, F., and J. Hovermale, 1968: "An Operational Six-Layer Primitive Equation Model," Journal of Applied Meteorology, Vol. 7, No. 4.

Stackpole, J., 1972: "The NMC 8-Layer Hemispheric Primitive Equation Model," to be published as an NMC Office Note.